

# Examination for UE3.78 « Brain-computer interfaces: from modeling to engineering » of master BME-BIN: Statistics Applied to Biology

Isabelle Rivals, Équipe de Statistique Appliquée, ESCPI Paris Tech

Solution

## 1 Thyroid cancer

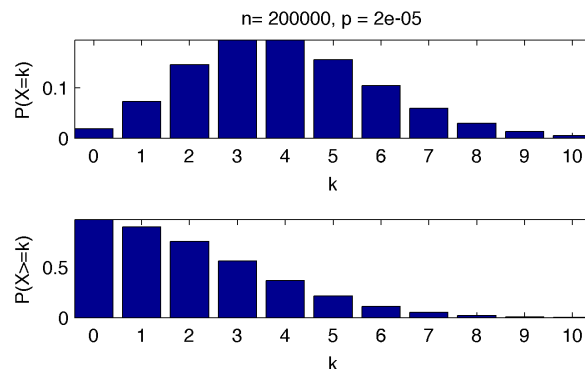
a)  $E(X) = n \times p_0$

b)  $X \sim Bi(n, p_0)$

c) Since  $n$  is large ( $> 20$ ) and  $p_0$  small ( $< 0.1$ ), the binomial law can be approximated with Poisson's law with parameter  $\lambda = n \times p_0 = 4$ . Using the table:

$$P(X \geq 10) = 1 - \sum_{k=0}^9 P(X = k) = 1 - 0.0183 - 0.0733 - 0.1465 - 0.1954 - 0.1954 - 0.1563 - 0.1042 - 0.0595 - 0.0298 - 0.0132 = 0.0081$$

d) Let  $H_0 : p = p_0 = 2 \cdot 10^{-5}$  and  $H_1 : p \neq p_0$ . If  $H_0$  is true,  $X \sim Bi(n, p_0)$ . Since  $P(X \geq 10 | p = p_0) = 0.0081 < \alpha/2 = 0.025$ ,  $H_0$  is rejected. We can state that the number of new cancers in the polluted region is abnormally high with a risk of 5% to be wrong.



## 2 Cardiovascular drug

a) Results of vartest2:

$$fstat = \frac{s_1^2}{s_2^2}, df1 = n_1 - 1, df2 = n_2 - 1$$

b) vartest2 performs Fisher's test of variance equality prior to Student's test, assuming Gaussian data.  $H_0 : \sigma_1^2 = \sigma_2^2$  versus  $H_1 : \sigma_1^2 \neq \sigma_2^2$ . Since  $p = 0.137 > 0.05$ ,  $H_0$  cannot be rejected. The variance can be considered equal in the two groups with an unknown risk of being wrong, and Student's test can be conducted.

c) Results of ttest2:

$$df = n_1 + n_2 - 2, sd = \sqrt{\frac{\sum_{k=1}^{n_1} (x_1^k - \bar{x}_1)^2 + \sum_{k=1}^{n_2} (x_2^k - \bar{x}_2)^2}{df}}, tstat = \frac{\bar{x}_1 - \bar{x}_2}{sd \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

- d) Since  $statse.tstat < 0$ ,  $\bar{x}_1 < \bar{x}_2$ . The heart beat rate is hence lower in group 1 (drug), which is the expected effect.
- e) ttest2 performs Student's test of mathematical expectation equality, assuming Gaussian data and equal variances.  $H_0 : \mu_1 = \mu_2$  versus  $H_1 : \mu_1 \neq \mu_2$ . Since  $p = 0.007 < 0.05$ ,  $H_0$  is rejected. The heart beat can be considered as significantly lower in the drug group, with a 5% risk to be wrong.
- f) It is hence meaningful to compute a c.i., which is given by Matlab :  $[-11.0483; -2.0073]$  (which does not include 0).